Perceptual video quality evaluation by means of physiological signals

Arnau González, Pablo; Althobaiti, Turke; Katsigiannis, Stamos; Ramzan, Naeem

Published in:
Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference on

DOI:
10.1109/QoMEX.2017.7965651

Published: 03/07/2017

Document Version
Peer reviewed version

Perceptual video quality evaluation by means of physiological signals

Pablo Arnau-González, Turke Althobaiti, Stamos Katsigiannis, Naeem Ramzan
University of the West of Scotland, School of Engineering and Computing
Paisley Campus, High St., Paisley, PA1 2BE, UK
E-mail: {Pablo.ArnauGonzalez, Turke.Althobaiti, Stamos.Katsigiannis, Naeem.Ramzan}@uws.ac.uk

***Abstract*—The proliferation of multimedia technology and its wide adoption by users has created the need for more effective metrics for Quality of Experience (QoE). Objective video quality metrics usually under-perform in terms of perceptual quality, thus evaluation is usually performed offline by people, an arduous and time consuming task that is also affected by external conditions and by user preferences. The use of physiological signals, recorded from users exposed to multimedia stimuli, has the potential to offer a more robust and unbiased method for evaluating perceptual quality. In this work, we propose the evaluation of the perceptual quality of video by means of cerebral (Electroencephalography - EEG) and peripheral (Electrocardiography - ECG and Electromyography - EMG) physiological signals. A machine learning approach is employed in order to map features extracted from these signals to a subjective video quality scale. Five 4K video sequences were encoded at different quality levels using the state-of-the-art HEVC codec and their quality was evaluated by real users while recording their physiological signals. The quality levels decided by the proposed model were then evaluated against the user-provided MOSs and the results demonstrated the potential of the proposed method for accurate perceptual video quality evaluation.**

Keywords*—*Quality of Experience; QoE; EEG; ECG; EMG; HEVC; perceptual video quality evaluation; physiological signals; 4K

I. Introduction

The extensive everyday use of multimedia technologies in recent years required efficient metrics for Quality of Experience (QoE) evaluation and helped advance the respective research field to a multidisciplinary field that studies the relations between various factors that affect quality, either human or system related. The widespread use of video streaming services requires an optimised balance between video quality and bandwidth requirements, thus being a critical aspect for any organisation that strives to provide the best experience to its customers within its planned technological and cost limitations. The industry's interest for efficient QoE methods led to numerous research works in the field of QoE evaluation, mainly focusing on audio [1] or visual [2] quality perception. Nevertheless, the progress in multimedia technology provided enhanced multimedia experiences that require more complex methods for assessing QoE and thus recent research focuses on multi-sensory approaches for efficiently handling QoE [3][4].

The usual target of QoE evaluation in multimedia streaming systems is the adjustment of video and audio quality settings in order to increase or decrease the quality and consequently increase or decrease the bandwidth requirements. This is usually achieved on the basis of objective quality metrics related to video and audio quality [5]. Nevertheless, objective video and audio quality metrics usually underperform in terms of perceptual quality and fail to efficiently capture the impact of the quality levels as perceived by human viewers or listeners. The most common way for assessing the perceptual quality of multimedia is by means of subjective evaluation, performed offline by people, an arduous and time consuming task that also has the drawback of being affected by external conditions during assessment and by user preferences.

The use of QoE assessment methods based on psychophysiology has been recently proposed [6] in order to address the limitations of the usual approaches in subjective quality evaluation. Psychophysiology approaches utilise the measurement of physiological signals in order to detect correlations to psychological responses in humans [6], [7]. Physiological signals are usually divided to signals originating from either the Central Nervous System (CNS) or the Autonomous Nervous System (ANS). The CNS category includes signals like the Electroencephalography (EEG) and Near-Infrared Spectroscopy (NIRS) signals, while the ANS category includes signals like the Electrocardiography (ECG), Electromyography (EMG) and Electrodermal Activity (EDA) signals. Many of these well-established signals have been used for affect recognition applications [8], [9], [10] and have the potential to be used for multimedia QoE assessment [6], [11].

Arndt et al. [12] proposed the use of brain activity (through EEG) and eye movement parameters in order to assess the quality of spatially degraded videos. Experiments using a consumer-grade EEG device (Emotiv EPOC [13]) and the SMI RED 5 remote eye tracker, showed that once a degraded area of the video was detected, subjects exhibited more focused attention on that area, as pupil diameter increases and the proportion of alpha activity decreases. According to Moldovan et al. [14], the main advantage of using EEG for QoE applications is that it enables the assessing of various QoE factors continuously over the duration of testing, without requiring the user to give any input about the visual quality as perceived by him/her. Their proposed QoE-EEG-Analyser attempted to automatically assess and quantify the impact of various factors contributing to user's QoE with multimedia services, using the participant's frustration level measured with a consumer-grade EEG system (Emotiv EPOC [13]). Preliminary results showed that frustration levels can indicate user's perceived QoE [14]. In another work, Perrin et al. [15] presented a novel multi-modal dataset for QoE analysis in emerging immersive

QoMEX2017 – Erfurt, Germany; 978-1-5386-4024-1/17/$31.00

multimedia applications. The dataset included recordings of physiological signals, such as EEG, ECG, and respiration, in order to evaluate the human experience while consuming immersive multimedia content in the form of audiovisual sequences. The analysis of the recorded signals in relation to the subjective ratings provided by the participants confirmed that the used modalities enable the distinction between low and high levels of immersiveness.

In this work, the authors propose the use of physiological data (EEG, ECG, and EMG) in order to evaluate the perceptual quality of 4K videos encoded using the state-of-the-art HEVC [16] compression standard [17], [18], [19]. The video sequences were encoded at different quality levels and their subjective quality was then evaluated by human subjects. Physiological signals were recorded for each subject during his/her exposure to the video stimuli. After an initial pre-processing stage, features were extracted from the recorded signals in order to train a machine learning model for the task of distinguishing between low and high visual quality. The experimental results are promising and show that physiological signal recordings can provide an indication of the visual quality perceived by the human visual system.

The rest of this paper is organised in three sections. Section II provides a detailed description of the proposed pre-processing, feature extraction and classification approach, as well as of the data acquisition procedure. The experimental results and discussion are provided in Section III, whereas conclusions are drawn in Section IV.

Fig. 1. Sample frames from the video sequences used in this study. Copyright and rights for each video sequence belong to the organisations/companies listed in Table I.

II. Methodology

A. Video stimuli

Five 4K video sequences, previously used for an evaluation of the HEVC video compression standard [20], were utilised in order to create the dataset of the video stimuli. The video sequences originated from different sources and as a result they have different spatio-temporal characteristics. The duration of each sequence is 10 s and their frame rate spans from 30 to 60 frames per second. All the sequences are in the $Y'C_BC_R$ colour space (ITU-R Rec. BT.709 [21]), with 8 bits per sample. Four compressed video sequences were created from each reference sequence by using the HEVC (HM-12.1, Main profile [22]) compression standard and four different fixed quantisation parameter (QP) settings. The QP settings were suitably selected in order for the subjective quality of the sequences to span a wide range of Mean Opinion Score (MOS) values. This procedure resulted to a total of 25 test sequences (including the reference sequences). Details about each sequence, including name, source, resolution, frame rate, bit rate, number of frames, and MOS retrieved from the tan et al. [20] study, are shown in Table I, while sample frames from each reference video sequence are shown on Fig. 1. Furthermore, the quality rating scale [23] used for establishing the MOS for the test sequences is shown in Table II.

B. Data collection

In order to examine the relation between physiological signals and perceived video quality, human subjects were used in order to subjectively rate the quality of each video sequence, while physiological signals were recorded during their exposure to the video stimuli. Twelve subjects in total participated in this proof-of-concept study, with their ages spanning between 23 and 35 years ($\mu = 30.42$, $\sigma = 3.40$). Moreover, all of them were holders of higher education degrees and non-experts in the field of image or video analysis and processing.

Each session consisted of placing the participant in a dark room sitting in front of a 4K screen (55 inch. Sony Bravia XD-93) at a distance of 1.07 m ($1.5 \times Height$ standard viewing distance for 4K video assessment according to ITU-R Rec. BT.2022 [24]). A thorough explanation of the experiment and the rating scale used was then given to the participant, who then proceeded to sign a consent form. After ensuring that the subject was sitting comfortably, the Emotiv EPOC [13] wireless EEG headset, the Shimmer wireless ECG [25] sensor and the Shimmer wireless EMG [25] sensor were attached to him/her. The 25 video sequences were then shown to the participant using the following procedure: The first reference sequence was shown followed by the four compressed versions in random order. After each sequence, the participant was asked to rate the perceived quality by clicking on the respective value of a rating scale (Table II). This procedure was then repeated until all video sequences were shown. A neutral video was also shown before each video sequence in order to return the subjects to a neutral stage.

Concerning the physiological data, EEG was recorded at a sampling rate of 128 Hz using 16 gold-plated contact-sensors that are fixed to flexible plastic arms of the Emotiv EPOC [13] wireless headset and are placed against the head in locations aligned with the following locations according to

TABLE I. TEST SEQUENCES AND THEIR PARAMETERS. BIT RATES AND MOS REFER TO THE BIT RATE AND MOS OF EACH OF THE FOUR ENCODED SEQUENCES

Sequence	Source/ Copyright	Resolution	FPS	Frames	Bit rate	MOS
BT709Birthday	Technicolor	3840 x 2160	50	500	[7.023, 3.701, 2.175, 1.321]	[8.75, 7.94, 7.16, 5.03]
Book	BBC	3840 x 2160	50	500	[6.123, 2.814, 1.662, 1.047]	[8.44, 7.91, 6.69, 4.78]
HomelessSleeping	Kamerawerk	3840 x 2160	60	600	[16.608, 5.526, 2.581, 1.488]	[8.28, 8.66, 8.00, 6.44]
Manege	4EVER	3840 x 2160	60	600	[17.840, 10.466, 6.139, 4.021]	[8.34, 7.66, 5.34, 3.50]
Traffic	Plannet, Inc.	4096 x 2048	30	300	[6.205, 3.137, 1.844, 1.056]	[8.43, 7.09, 5.50, 3.50]

TABLE II. SUBJECTIVE QUALITY SCALE USED FOR ESTABLISHING THE MOS FOR THE VIDEO SEQUENCES EVALUATED IN THIS STUDY

Score	Description of scores
10	Denotes a quality of reproduction that is perfectly faithful to the original. No further improvement is possible
9 8	If some difference is "seen or even only thought to be seen" (8 to be used when this happens for more than one part of this image)
7 6	If the viewer is sure to have seen some difference (6 to be used when this happens for more than one part of the image)
5 4	When the differences are evident and visible with no particular effort (4 to be used when this happens for more than one par of the image)
3 2	When differences are many and annoying (2 to be used when this is severely impairing the image)
1 0	When the image is severely impaired and very far from the original (0 to be used when this happens all over the image making it hardly understandable). Score of 0 denotes a quality of reproduction that has no similarity to the original. A worse quality cannot be imagined.

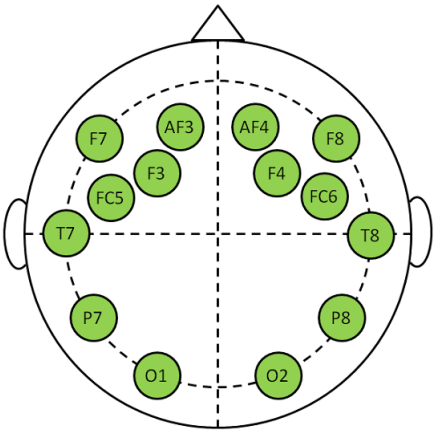

Fig. 2. Positioning of the Emotiv EPOC contact-sensors according to the International 10-20 system [26]

the International 10-20 system: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, M1 and M2, as shown on Fig. 2. ECG and EMG signals were recorded at 256 Hz, with the ECG using four standard electrodes placed on both lower ribs and upper clavicle, and the EMG using three standard electrodes placed on the upper trapezius muscles, as shown on Fig. 3.

It must be noted that approval for conducting this study and for publishing anonymised results was given by the University of the West of Scotland's *University Ethics Committee*.

TABLE III. EXTRACTED FEATURES FROM EACH MODALITY

Modality	Extracted features	# Features
EEG	Logarithm of the PSD for the alpha, beta, gamma and theta bands of each of the 14 electrodes	56
ECG	Maxima, minima, mean, media, standard deviation and range from the raw signal and the derivative of PQ, QS and ST complexes. Number of intervals with latency > 50 ms from HRV. PSD from HRV between the intervals $[0\,,0.2]$, $[0.2\,,0.4]$, $[0.4\,,0.6]$ and $[0.6\,,0.8]$. Maxima, minima, mean, median, standard deviation and range from HRV histogram.	84
EMG	Maxima, minima, median, mean, standard deviation, and number of times per time unit that the signal reached both the minima and the maxima, extracted from the a) raw signal, b) first derivative, and c) second derivative	21

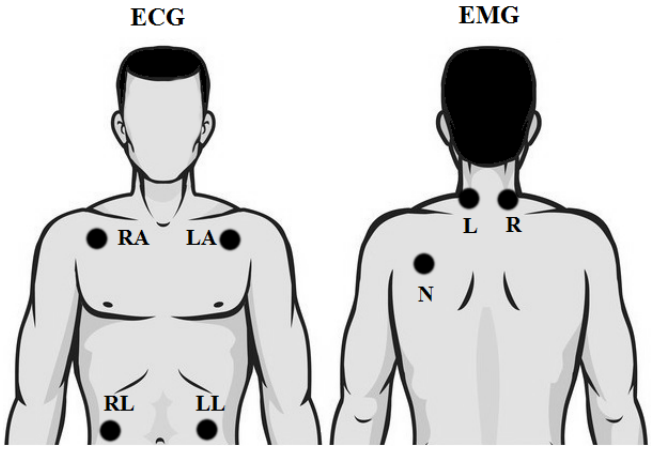

Fig. 3. Electrode positioning for the ECG and EMG sensors

C. Pre-processing

Pre-processing is an essential step in the process of physiological signal analysis due to the extremely noisy nature of the signals. The EMG signals have been pre-processed using the Augsburg Biosignal Toolbox (AuBT) [27]. Pre-processing consisted of first cropping all the samples with an amplitude inside the highest or the lowest 3% of amplitude values, and then by applying a lowpass filter with 0.4 Hz as cutoff frequency. EEG signals were filtered and then cleaned from artefacts (without channel rejection) using the EEGLab [28] specialized toolbox. Pre-processing consisted of filtering using a butterworth bandpass filter between $4-64$ Hz and of artefact cleaning by removing flatline windows and low frequency drifts, and by averaging for the noisy channels. Contrary to the EEG and EMG signals, ECG signals are less susceptible to interferences due to their higher voltage amplitudes and thus require no further processing.

D. Feature extraction

After the pre-processing of the recorded signals, the following features were extracted from each signal in order to be used for the classification stage:

1) EEG-based features: The power spectral densities (PSD) of different frequency bands have been commonly utilised to describe patterns in EEG signals [29], [9], [30], [8]. These features are computed using the Welch estimate of spectral power and by averaging across the components belonging to the frequency band. In this work, PSD features are computed from the theta (θ: 4 *Hz* - 8 *Hz*), alpha (α: 8 *Hz* - 13 *Hz*), beta (β: 13 *Hz* - 30 *Hz*) and gamma (γ: 30 *Hz* - 64 *Hz*) frequency bands of each of the 14 channels of the recorded EEG signals. The logarithm of the PSD is then used as a feature, leading to a total of 56 EEG-based features (4 for each of the 14 channels). The feature vector F_{EEG} of the EEG-based features is defined as follows: Let $F_{i\theta}$, $F_{i\alpha}$, $F_{i\beta}$, and $F_{i\gamma}$ be the logarithm of the PSD for the signal of the i-th electrode, $i = 1, 2, ..., 14$, for the theta, alpha, beta, and gamma bands respectively. Then $F_{EEG} = [\, F_{1\theta} \quad F_{1\alpha} \quad F_{1\beta} \quad F_{1\gamma} \quad ... \quad F_{14\theta} \quad F_{14\alpha} \quad F_{14\beta} \quad F_{14\gamma} \,]$.

2) ECG-based features: Different kinds of features have been extracted from the ECG signal. First, the PQRST complexes of the ECG signal are detected using the Augsburg Biosignal Toolbox (AuBT) [27]. Then PQ, QS and ST complexes are extracted by subtracting the time elapsed for the corresponding pair of peaks of each of the signals. For each of these complexes, statistical features are extracted from the raw signal, as well as from its first derivative. Furthermore, some of the most consistently utilised ECG-based features are the heart rate (HR) and heart rate variability (HRV) specific parameters in the time and frequency domain respectively [9], [8]. Consequently, in addition to the features mentioned above, HRV features were also computed and consisted of the number of intervals with a latency greater than $50\ ms$, statistical features from the histogram of HRV, and the power spectral density between the intervals $[0\ , 0.2]$, $[0.2\ , 0.4]$, $[0.4\ , 0.6]$ and $[0.6\ , 0.8]$. The statistical features from the histogram of HRV consisted of maxima and minima, mean and median, standard deviation and the range of the signal. The feature vector is then formed by concatenating all these features: $F_{ECG} = [F_{stats} \quad F_{HRV}]$, where F_{stats} represents the concatenation of all the statistical features computed on the raw ECG signal, and F_{HRV} represents the feature vector created by the concatenation of all the HRV-related features extracted. This procedure resulted in a total of 84 ECG-based features.

3) EMG-based features: A set of statistical features were extracted from the EMG signal, as well as from its first and its second derivative, using the Augsburg Biosignal Toolbox (AuBT) [27]. These statistical features were the mean, median, standard deviation, minima, maxima, and the number of times per time unit that the signal reached both the minima and the maxima. Finally, the feature vector F_{EMG} was created by the concatenation of all these features: $F_{EMG} = \left[F_{emg} \quad F_{\frac{\delta emg}{\delta t}} \quad F_{\frac{\delta^2 emg}{\delta^2 t}} \right]$, leading to a total of 21 EMG-based features.

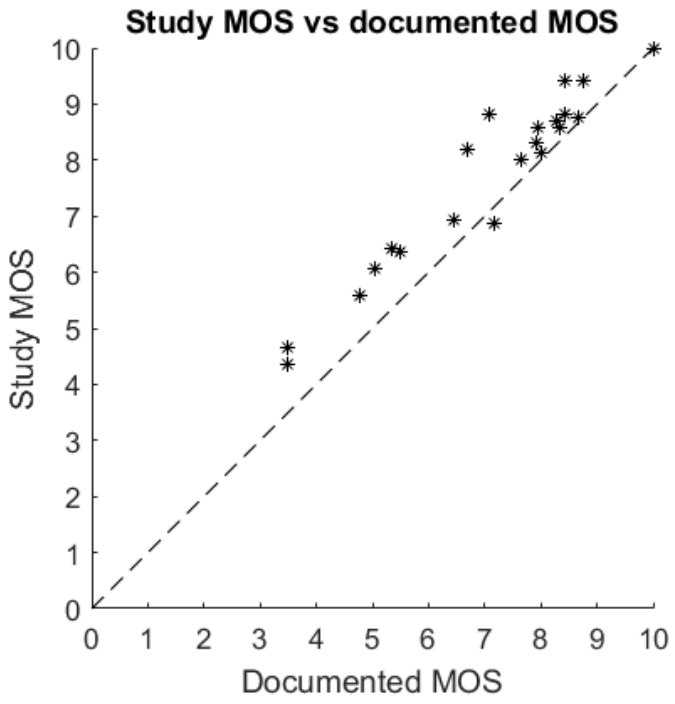

Fig. 4. The MOS reported in this study compared to the MOS reported in [20] for the same video sequences

III. Experimental evaluation

A. Evaluation of captured data in terms of MOS

In order to establish the quality of the captured data, the average MOS reported by the participants of this study for each video sequence was compared to the average MOS reported for the same video sequences in [20]. It is critical to establish whether the current ratings agree with the already available ratings since if the captured MOSs differ significantly, then the findings of this study would be disputable due to unsuitable data. The Pearson's Correlation Coefficient (PCC) was computed in order to evaluate their similarity, resulting to a very strong correlation (PCC $= 0.9726$) between the MOS from [20] and this study's MOS. Furthermore, a one-way analysis of variance (ANOVA) showed that there is no statistically significant difference ($p = 0.3053$) between the ratings. The similarity in ratings is also evident on Fig. 4, where the MOS reported in this study is plotted against the MOS from [20].

B. Classification experiments

Supervised classification experiments were conducted in order to assess the suitability of the physiological signals recorded for mapping physiological responses to the visual quality of video sequences as perceived by human viewers. Since the subjects were first shown the original reference video sequence, in this experimental evaluation we examine whether the perception of low or high quality compared to the original is expressed in the physiological recordings. The values of the rating scale used for the subjective quality assessment ranged between 0 and 10. The ratings provided by the participants were then labeled as *Low Quality* or *High Quality* by applying a threshold as follows: If the subject reported a score of 9 or 10, meaning that according to the scale (Table II), the subject

TABLE IV. CLASSIFICATION RESULTS FOR ALL THE MODALITIES AND CLASSIFICATION ALGORITHMS EXAMINED

	1-NN		3-NN		SVM-LINEAR		SVM-RBF	
	Accuracy (St.D.)	F-score	Accuracy (St.D.)	F-score	Accuracy (St.D.)	F-score	Accuracy (St.D.)	F-score
EEG	0.5932 (0.089)$^{+}$	0.5844	0.5532(0.073)	0.5439	0.4806 (0.103)	0.4751	0.5533 (0.137)	0.3949
ECG	0.5247 (0.113)	0.5310	0.5250 (0.100)	0.5238	0.5426 (0.121)	**0.5527**	0.5533 (0.137)	0.3949
EMG	0.4861 (0.056)*	0.4844	0.4919 (0.068)	0.4898	0.5532 (0.136)	0.3949	**0.5839 (0.060)**$^{+}$	**0.5722**
EEG+ECG	0.5208 (0.107)	0.5286	0.5220 (0.101)	0.5213	0.4769 (0.111)	0.4797	0.5533 (0.137)	0.3949
EEG+EMG	0.6071 (0.085)$^{*+}$	0.5982	**0.5627 (0.080)**$^{+}$	0.5525	0.4670 (0.111)	0.4600	0.5533 (0.137)	0.3949
EEG+ECG+EMG	0.5208 (0.107)	0.5285	0.5220 (0.101)	0.5213	0.4798 (0.114)	0.4827	0.5533 (0.137)	0.3949
Feature selection	**0.6455 (0.111)**$^{*+}$	**0.6388**	**0.5636 (0.137)**	**0.5658**	**0.5900 (0.102)**	0.5001	0.5726 (0.120)	0.5545

Note: * indicates a statistically significant difference ($p < 0.05$) when compared to classifying according to the class ratio, while $^{+}$ indicates a statistically significant difference ($p < 0.05$) when compared to the random classifier. Values in bold indicate the highest value achieved for each classifier.

did not perceive any degradation in the quality of the video sequence, then the label was set to *High Quality*. On the other hand, if the subject reported a score of 8 or lower, meaning that the subject perceived noticeable degradation in the quality of the video sequence, then the label was set to *Low Quality*. The reason for mapping the recorded ratings to *High* and *Low Quality* is based on studies that showed that there is correlation between EEG/physiological signals and visual quality levels [6], [12].

Two classification algorithms were employed in order to evaluate the efficiency of the proposed method, namely the k-Nearest Neighbour (k-NN) for $k = 1$ and $k = 3$ and Support Vector Machines (SVM), using the Linear version, as well as the Radial Basis Function (RBF) kernel. For the validation of the classification performance and in order to avoid over-fitting, a *leave-one-subject-out* cross validation scheme was applied, i.e. at each fold of the cross validation, all the samples of one subject were used for testing while all the other samples were used for training. Seven different feature sets were evaluated using the aforementioned classification methods. The classification performance achieved for the features extracted from each physiological signal modality was examined, along with the classification performance for the fusion of EEG+ECG, EEG+EMG and EEG+ECG+EMG -based features. Moreover, the Sequential Feature Selection (SFS) algorithm was applied in order to select the most descriptive features amongst all the computed features (EEG+ECG+EMG) and the performance of the selected feature subset was also evaluated. At the feature selection stage, the cross-validated classification accuracy was used as the criterion function for selecting the optimal feature subset.

Results are reported in Table IV in terms of classification accuracy and the weighted F-score. It must be noted that since the ECG recordings contained two channels, only the results for the best performing channel are reported. From Table IV it is evident that the best classification accuracy and F-score among individual modalities is achieved for the EEG-based features using the 1-NN classifier (0.5932 and 0.5844 for accuracy and F-score respectively), while the second best performance was achieved for the ECG-based features. Taking into consideration the fusion approaches, the best performance was achieved for the fusion of the EEG and EMG features using the 1-NN classifier (0.6071 accuracy and 0.5982 F-score). Nevertheless, the best overall performance was achieved for the feature selection approach (0.6455 accuracy and 0.6388 F-score), using a feature subset containing a total of 5 features out of the initial 161, with the fusion of EEG-EMG providing the second best overall performance when used in combination with the 1-NN classifier.

In order to validate the significance of the classification results, the achieved accuracies were compared against the analytically computed accuracies for a random classifier (Accuracy $= 0.50$) and a classifier that votes according to the class ratio (F-score $= 0.50$). Since we cannot assure the normality and homoscedasticity of the data, which makes the traditional t-test ineligible, p-values have been computed using a non-parametric alternative to Student's t-test, i.e. the Wilcoxon's signed rank test. As shown in Table IV, the results for the feature selection approach using the 1-NN classifier, as well as the results for the EEG+EMG fusion approach using the same classifier, have a statistically significant difference compared to the results for random voting ($p < 0.05$) and voting according to the class ratio ($p < 0.05$).

The use of the SVM-Linear classifier did not produce any statistically significant results for any modality, whereas for SVM-RBF only the results for the EMG-based features were statistically significant when compared to classifying according to the class ratio but not for random voting. Similarly, the use of 3-NN provided statistically significant results only for the fusion of EEG+EMG features when compared to classifying according to the class ratio but not for random voting. Finally, for 1-NN, the results for EEG-based features were statistically significant when compared to classifying according to the class ratio but again not for random voting, while EMG-based features provided statistically significant results only against random voting. Consequently, only the results for the feature selection approach and the EEG+EMG fusion approach can be considered as reliable. As a result, it can be argued that the features extracted individually from each modality are not descriptive enough for distinguishing between the perception of lower or higher visual quality in the case of video stimuli. Nevertheless, the success of the feature selection approach validates that there is correlation between the examined physiological signals and the perception of visual quality. Furthermore, the use of only 5 out of 161 features provides an indication that the feature space is highly redundant and noisy.

IV. CONCLUSION

In this work, the authors proposed a method for assessing the perceptual visual quality of video sequences through the

use of physiological signals. Features extracted from EEG, ECG, and EMG signals were evaluated through different classification algorithms for the task of classifying the perceived quality of a video sequence as low or high, compared to the original reference video sequence. The experimental evaluation provided promising results, achieving a classification accuracy of 0.6455 and an F-score of 0.6388 when using the feature selection approach. Furthermore, among the individual modalities, the EEG provided enhanced performance compared to ECG and EMG -based features, achieving a classification accuracy of 0.5932 and an F-score of 0.5844, whereas the combination of EEG and EMG provided the highest performance among individual modalities and fusion approaches, while achieving the second best overall performance (0.6071 accuracy and 0.5982 F-score). These results demonstrate the ability of the examined physiological signals to encode information about the perceived visual quality and provide evidence for the feasibility of using physiological signals for perceptual visual quality evaluation.

ACKNOWLEDGMENT

The authors would like to thank the people that participated in the experiments for their help in realising this work.

REFERENCES

[1] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 2, 2001, pp. 749–752.

[2] S. Winkler, "Digital video quality: vision models and metrics," 2005.

[3] S. Möller and A. Raake, "Digital video quality: vision models and metrics," 2014.

[4] G. Ghinea, C. Timmerer, W. Lin, and S. R. Gulliver, "Mulsemedia: State of the art, perspectives, and challenges," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 11, no. 1s, pp. 17:1–17:23, Oct. 2014.

[5] K. Kunze and D. Strohmeier, "Examining subjective evaluation methods used in multimedia quality of experience research," in *2012 Fourth International Workshop on Quality of Multimedia Experience*, July 2012, pp. 51–56.

[6] U. Engelke, D. P. Darcy, G. H. Mulliken, S. Bosse, M. G. Martini, S. Arndt, J. N. Antons, K. Y. Chan, N. Ramzan, and K. Brunnstrom, "Psychophysiology-based qoe assessment: A survey," *IEEE Journal of Selected Topics in Signal Processing*, vol. PP, no. 99, pp. 1–1, 2016.

[7] P. Arnau-Gonzlez, M. Arevalillo-Herrez, and N. Ramzan, "Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals," *Neurocomputing*, vol. 244, pp. 81 – 89, 2017.

[8] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis ;using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

[9] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, Jan 2012.

[10] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, 2017.

[11] D. Senkowski, T. R. Schneider, J. J. Foxe, and A. K. Engel, "Crossmodal binding through neural coherence: implications for multisensory processing," *Trends Neurosci.*, vol. 31, no. 8, pp. 401–409, Aug 2008.

[12] S. Arndt, J. Radun, J. N. Antons, and S. Móller, "Using eye-tracking and correlates of brain activity to predict quality scores," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, Sept 2014, pp. 281–285.

[13] N. A. Badcock, P. Mousikou, Y. Mahajan, P. de Lissa, J. Thie, and G. McArthur, "Validation of the Emotiv EPOC EEG gaming system for measuring research quality auditory ERPs," *PeerJ*, vol. 1, p. e38, Feb. 2013.

[14] A. N. Moldovan, I. Ghergulescu, S. Weibelzahl, and C. H. Muntean, "User-centered eeg-based multimedia quality assessment," in *2013 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, June 2013, pp. 1–8.

[15] A.-F. N. M. Perrin, H. Xu, E. Kroupi, M. Řeřábek, and T. Ebrahimi, "Multimodal dataset for assessment of quality of experience in immersive multimedia," in *Proceedings of the 23rd ACM International Conference on Multimedia*, ser. MM '15. New York, NY, USA: ACM, 2015, pp. 1007–1010.

[16] "H.265: High efficiency video coding," ITU-T Rec. H.265, December 2016.

[17] N. Ramzan, Z. Pervez, and A. Amira, "Quality of experience evaluation of h.265/mpeg-hevc and vp9 comparison efficiency," in *2014 26th International Conference on Microelectronics (ICM)*, Dec 2014, pp. 220–223.

[18] T. K. Tan, M. Mrak, V. Baroncini, and N. Ramzan, "HEVC verification test results," JCTVC-Q0204, April 2014.

[19] V. B. T. K. Tan, M. Mrak and N. Ramzan, "Report on HEVC compression performance verification testing," Joint Collab. Team Video Coding (JCT-VC), 2014.

[20] T. K. Tan, R. Weerakkody, M. Mrak, N. Ramzan, V. Baroncini, J. R. Ohm, and G. J. Sullivan, "Video quality evaluation methodology and verification testing of hevc compression performance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 76–90, Jan 2016.

[21] "Parameter values for the HDTV standards for production and international programme exchange," ITU-R Rec. BT.709, April 2002.

[22] JCT-VC of ITU-T SG16/Q6 and ISO/IEC JTC1/SC29/WG11, "HEVC HM Reference Software Codebase, Version 12.1," October 2013. [Online]. Available: http://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

[23] "Subjective video quality assessment methods for multimedia applications," ITU-T Rec. P.910, April 2008.

[24] "General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays," ITU-R Rec. BT.2022, August 2012.

[25] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, and V. Cionca, "SHIMMER A Wireless Sensor Platform for Noninvasive Biomedical Research," *IEEE Sensors Journal*, vol. 10, pp. 1527–1534, 2010.

[26] D. Wijayasekara and M. Manic, "Human machine interaction via brain activity monitoring," in *2013 6th International Conference on Human System Interactions (HSI)*, June 2013, pp. 103–109.

[27] J. Wagner, "Augsburg Biosignal Toolbox (AuBT)," University of Augsburg, 2005.

[28] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9 – 21, 2004.

[29] R. J. Davidson, "Affective neuroscience and psychophysiology: toward a synthesis," *Psychophysiology*, vol. 40, no. 5, pp. 655–665, Sep 2003.

[30] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "DECAF: MEG-based multimodal database for decoding affective physiological responses," *IEEE Transactions on Affective Computing*, vol. 6, no. 3, pp. 209–222, July 2015.